Contents

1 Introduction 1
1.1 General Background 1
1.2 Focus of This Book 4
1.3 Organization of the Remainder of the Book 10
References 11

2 Complex Networks 15
2.1 Basic Concepts of Graphs 15
2.1.1 Graph Definitions 16
2.1.2 Connectivity 20
2.1.3 Paths and Cycles 24
2.1.4 Subgraphs 26
2.1.5 Trees and Forest 28
2.1.6 Graph Representation 29
2.2 Complex Network Models 31
2.2.1 Random Networks 32
2.2.2 Small-World Networks 34
2.2.3 Scale-Free Networks 35
2.2.4 Random Clustered Networks 37
2.2.5 Core-Periphery Networks 37
2.3 Complex Network Measures 40
2.3.1 Degree and Degree-Correlation Measures 40
2.3.2 Distance and Path Measures 43
2.3.3 Structural Measures 44
2.3.4 Centrality Measures 47
2.3.5 Classification of the Network Measurements 54
2.4 Dynamical Processes in Complex Networks 55
2.4.1 Random Walks 55
2.4.2 Lazy Random Walks 62
2.4.3 Self-Avoiding Walks 63

2.4.4 Tourist Walks 63
2.4.5 Epidemic Spreading 65
2.5 Chapter Remarks 67
References 67

3 Machine Learning 71
3.1 Overview of Machine Learning 71
3.2 Supervised Learning 74
3.2.1 Mathematical Formalization and Fundamental Assumptions 74
3.2.2 Overview of the Techniques 77
3.3 Unsupervised Learning 78
3.3.1 Mathematical Formalization and Fundamental Assumptions 78
3.3.2 Overview of the Techniques 80
3.4 Semi-Supervised Learning 82
3.4.1 Motivations 82
3.4.2 Mathematical Formalization and Fundamental Assumptions 83
3.4.3 Overview of the Techniques 85
3.5 Overview of Network-Based Machine Learning 86
3.6 Chapter Remarks 87
References 88

4 Network Construction Techniques 93
4.1 Introduction 93
4.2 Similarity and Dissimilarity Functions 96
4.2.1 Formal Definitions 96
4.2.2 Examples of Vector-Based Similarity Functions 98
4.3 Transforming Vector-Based Data into Networks 104
4.3.1 Analysis of k-Nearest Neighbors and ϵ-Radius Networks 106
4.3.2 Combination of k-Nearest Neighbors and ϵ-Radius Network Formation Techniques 108
4.3.3 b-Matching Networks 109
4.3.4 Linear Neighborhood Networks 110
4.3.5 Relaxed Linear Neighborhood Networks 112
4.3.6 Network Formation Using Clustering Heuristics 114
4.3.7 Network Formation Using Overlapping Histogram Segments 115
4.3.8 More Advanced Network Formation Techniques 119
4.4 Transforming Time Series Data into Networks 121
4.4.1 Cycle Networks 124
4.4.2 Correlation Networks 125
4.4.3 Recurrence Networks 126
4.4.4 Transition Networks 126

4.5 Classification of Network Formation Techniques 127
4.6 Challenges in Transforming Unstructured Data to Networked Data 128
4.7 Chapter Remarks 130
References 130

5 Network-Based Supervised Learning 133
5.1 Introduction 133
5.2 Representative Network-Based Supervised Learning Techniques . 135
5.2.1 Classification Using *k*-Associated Graphs 136
5.2.2 Network Learning Toolkit (NetKit) 137
5.2.3 Classification Using *Ease of Access* Heuristic 138
5.3 Chapter Remarks 140
References 141

6 Network-Based Unsupervised Learning 143
6.1 Introduction 143
6.2 Community Detection 146
6.2.1 Relevant Concepts and Motivations 146
6.2.2 Mathematical Formalization and Fundamental Assumptions 148
6.2.3 Overview of the State-of-the-Art Techniques 150
6.2.4 Community Detection Benchmarks 150
6.3 Representative Network-Based Unsupervised Learning Techniques 151
6.3.1 Betweenness 152
6.3.2 Modularity Maximization 153
6.3.3 Spectral Bisection Method 157
6.3.4 Community Detection Using Particle Competition 159
6.3.5 Chameleon 161
6.3.6 Community Detection by Space Transformation and Swarm Dynamics 163
6.3.7 Synchronization Methods 167
6.3.8 Finding Overlapping Communities 169
6.3.9 Network Embedding and Dimension Reduction 174
6.4 Chapter Remarks 176
References 177

7 Network-Based Semi-Supervised Learning 181
7.1 Introduction 181
7.2 Network-Based Semi-Supervised Learning Assumptions 183
7.3 Representative Network-Based Semi-Supervised Learning Techniques 185
7.3.1 Maximum Flow and Minimum Cut 186
7.3.2 Gaussian Field and Harmonic Function 187
7.3.3 Tikhonov Regularization Framework 189

7.3.4 Local and Global Consistency 190
7.3.5 Adsorption 191
7.3.6 Semi-Supervised Modularity Method 194
7.3.7 Interaction Forces 197
7.3.8 Discriminative Walks (D-Walks) 198
7.4 Chapter Remarks 202
References 203

8 Case Study of Network-Based Supervised Learning: High-Level Data Classification 207
8.1 A Quick Overview of the Chapter 208
8.2 Motivation 209
8.3 Model Description 212
8.3.1 Fundamental Ideas Behind the Model 212
8.3.2 Derivation of the Hybrid Classification Framework 216
8.4 Possible Ways of Composing High-Level Classifiers 219
8.4.1 High-Level Classification Using a Mixture of Complex Network Measures 219
8.4.2 High-Level Classification Using Tourist Walks 222
8.5 Numerical Analysis of the High-Level Classification 226
8.5.1 An Illustrative Example 226
8.5.2 Parameter Sensitivity Analysis 227
8.6 Application: Handwritten Digits Recognition 231
8.6.1 Motivation 231
8.6.2 Description of the MNIST Data Set 232
8.6.3 A Suitable Similarity Measure for Images 232
8.6.4 Configurations of the Low-Level Classification Techniques 233
8.6.5 Experimental Results 233
8.6.6 Illustrative Examples: High-Level Classification vs. Low-Level Classification 234
8.7 Chapter Remarks 238
References 238

9 Case Study of Network-Based Unsupervised Learning: Stochastic Competitive Learning in Networks 241
9.1 A Quick Overview of the Chapter 241
9.2 Description of the Stochastic Competitive Model 242
9.2.1 Intuition of the Model 243
9.2.2 Derivation of the Transition Matrix 244
9.2.3 Definition of the Stochastic Nonlinear Dynamical System 252
9.2.4 Method for Estimating the Number of Communities 254
9.2.5 Method for Detecting Overlapping Structures 255
9.2.6 Parameter Sensitivity Analysis 255
9.2.7 Convergence Analysis 259

9.3 Theoretical Analysis of the Model 263
9.3.1 Mathematical Analysis 263
9.3.2 Linking the Particle Competition Model and the Classical Multiple Independent Random Walks System 274
9.3.3 A Numerical Example 276
9.4 Numerical Analysis of the Detection of Overlapping Vertices and Communities 280
9.4.1 Zachary's Karate Club Network 280
9.4.2 Dolphin Social Network 282
9.4.3 Les misérables Novel Network 283
9.5 Application: Handwritten Digits and Letters Clustering 284
9.5.1 Brief Information of the Handwritten Digits and Letters Data Sets 284
9.5.2 Determining the Optimal Number of Particles and Clusters 285
9.5.3 Handwritten Data Clustering 285
9.6 Chapter Remarks 288
References 289

10 Case Study of Network-Based Semi-Supervised Learning: Stochastic Competitive-Cooperative Learning in Networks 291
10.1 A Quick Overview of the Chapter 291
10.2 Description of the Stochastic Competitive-Cooperative Model 292
10.2.1 Differences of the Semi-Supervised and the Unsupervised Versions 293
10.2.2 Familiarizing with the Semi-Supervised Environment 295
10.2.3 Deriving the Modified Competitive Transition Matrix 295
10.2.4 Modified Initial Conditions of the System 296
10.3 Theoretical Analysis of the Model 298
10.3.1 Mathematical Analysis 298
10.3.2 A Numerical Example 301
10.4 Numerical Analysis of the Model 304
10.4.1 Simulation on a Synthetic Data Set 304
10.4.2 Simulations on Real-World Data Sets 306
10.5 Application: Detection and Prevention of Error Propagation in Imperfect Learning 308
10.5.1 Motivation 308
10.5.2 Detecting Imperfect Training Data 309
10.5.3 Preventing Label Propagation from Imperfect Training Data 311
10.5.4 Definition of the Modified Learning System to Withstand Imperfect Data 313

10.5.5 Parameter Sensitivity Analysis 313
10.5.6 Computer Simulations 317
10.6 Chapter Remarks 320
References 321

Index 323

http://www.springer.com/978-3-319-17289-7

Machine Learning in Complex Networks

Christiano Silva, T.; Zhao, L.
2016, XVIII, 331 p. 87 illus., 80 illus. in color., Hardcover
ISBN: 978-3-319-17289-7